

METHOD, SYSTEM, AND PROGRAM FOR PROVIDING DATA TO AN
APPLICATION PROGRAM FROM A FILE IN A FILE SYSTEM

BACKGROUND OF THE INVENTION

5 1. Field of the Invention

[0001] The present invention relates to a method, system, and program for providing data to an application program from a file in a file system.

2. Description of the Related Art

10 [0002] In hierarchical storage management (HSM) systems, files are migrated to tape storage when the data stored in disk cache reaches a certain threshold. HSM systems migrate files to tape to make room for further files being used in the system. If an application program requires data from a file migrated to tape to continue processing, then the application processing may be delayed while the entire file is staged back onto
15 tape. Such HSM systems optimize the use of the storage space because less frequently used files may be migrated to tape and removed from the disk storage to make space available for new data, thereby increasing the effective storage capacity of the disk storage.

[0003] The delay in retrieving a migrated file from the tape storage is significant for
20 database programs. The data in a database is stored in tables. Each database table is comprised of one or more tablespaces where the table data is stored. Each tablespace is comprised of one or more separate data files. In the prior art, all the tablespace files typically remain on the disk cache if the database tables which they comprise are open for writes by the database program. Such tablespace files remain on the disk space even if
25 they have not been accessed for a long time. Thus, the benefits of hierarchical storage management (HSM) that would be realized by releasing old archived tablespace files cannot be realized when a substantial portion of the disk space is used to store tablespace files of an open database.

00894078-062301

[0004] For this reason, there is a need in the art to provide improved techniques for managing files in a file system.

SUMMARY OF THE PREFERRED EMBODIMENTS

- 5 [0005] Provided is a method, system, and program for managing files in a file system. A plurality of files are provided in a primary storage used by an application program. A criteria is applied to determine files to release in the primary storage that have been copied to a secondary storage. A request is received for data from the application program in one file that was released and resides on the secondary storage. Data is read
- 10 from the requested file in the secondary storage into a memory accessible to the application program. Data is provided from the file in the memory to the application program before the entire file has been read from the secondary storage into the memory.
- [0006] In further implementations, the primary storage stores a partial version of at least one released file, wherein the partial version includes a portion of the data in at least one
- 15 released file.
- [0007] Still further, a stage attribute is associated with each file indicating whether to stage the file transferred from the secondary storage to the memory into the primary storage. Data for the file is staged to the primary storage from the memory that was transferred from the secondary storage only if the stage attribute indicates that data from
- 20 the file is to be staged.
- [0008] In further implementations, groups of component files are accessed by the application program. In such case, the primary storage maintains a partial version of each released component file included in one of the groups accessed by the application program, wherein the partial version includes a portion of the released component file.
- 25 [0009] In certain implementations, the application program comprises a database program and the groups of component files comprise tablespaces that the database program has opened, wherein the component files of one opened tablespace are eligible for release according to the criteria.

09094078-062801

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] Referring now to the drawings in which like reference numbers represent corresponding parts throughout:

FIG. 1 is an illustration of a computing environment in which aspects of the invention are implemented;

FIG. 2 illustrates a data structure for metadata in accordance with implementations of the invention;

FIGs. 3 and 4 illustrate files of a tablespace stored in the file system in accordance with implementations of the invention; and

FIGs. 5a and 5b illustrate logic to I/O requests to files in the file system in accordance with implementations of the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0011] In the following description, reference is made to the accompanying drawings which form a part hereof and which illustrate several embodiments of the present invention. It is understood that other embodiments may be utilized and structural and operational changes may be made without departing from the scope of the present invention.

[0012] FIG. 1 illustrates a computing environment implementation of the invention. A computer 2, which may comprise any computing device known in the art, including a desktop computer, mainframe, workstation, personal computer, hand held computer, palm computer, laptop computer, telephony device, network appliance, etc., includes a file system 4 and an application program 8. The file system 4 may comprise any file system that an operating system provides to organize and manage files known in the art, such as the file system used with the Sun Microsystems Solaris operating system, Unix file system or any other file system known in the art.** The application program 8 may comprise any application known in the art that creates and accesses data files in the file system 4, such as a database program, word processing program, software development

09894078 062801

tool or any other application program known in the art. A network 18, which may comprise any network system known in the art, such as Fibre Channel, Local Area Network (LAN), an Intranet, Wide Area Network (WAN), Storage Area Network (SAN), etc., enables communication between the computer 2, primary storage 10, and secondary storage 12. Alternatively, the computer 2 may be connected to the disk cache 10 and tape library 12 via direct transmission lines or cables (not shown). Data transferred between the primary storage, such as a disk cache, 10 and the secondary storage 12, such as a tape library, transferred through the file system 4 in the computer 2 or, alternatively, directly between the primary storage 10 and secondary storage 12 via the network 18 or a direct transmission line (not shown).

[0013] In the described implementations, the file system 4 further includes programs for managing the storage of files in the file system 4 in a primary storage 10 and secondary storage 12. In certain implementations, the primary storage 10 comprises a disk cache or group of interconnected hard disk drives that implement a single storage space. The applications 8 process data stored in the primary storage 10. The secondary storage 12 is used for maintaining a backup copy of files in the file system 4 and for expanding the overall available storage space. In certain implementations, the secondary storage 12 comprises a slower access and less expensive storage system than the primary storage 12. For instance, the secondary storage 12 may comprise a tape library including one or more tape drives and numerous tape cartridges, an optical library, slower and less expensive disk drives, etc. In certain implementations, once a tape cartridge is mounted in a tape drive, data may be transferred between the primary 10 and secondary 12 storage.

[0014] In certain implementations, the file system 4 is capable of performing Hierarchical Storage Management (HSM) related functions, such as automatically archiving files in the primary storage 10 in the secondary storage 12. Files are archived when they meet a set of archive criteria, such as age, file size, time last accessed, etc. The file system 4 may also perform staging operations to copy data archived on the secondary storage 12 to the primary storage 10 to make available to the applications 8.

09894078 "062301

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225

Partial Release 56: indicates that the first n bytes of the file are maintained in the primary storage 10 after the release operation, where n may be a user settable parameter..

Segment 58: indicates that the file data is stored in separate segments as described in the copending and commonly assigned patent application entitled "Method, System, and Program for Managing Files in a File System," having attorney docket no. P6433 filed on the same date herewith, which patent application is incorporated herein by reference in its entirety.

Offline 60: indicates that the file is currently resident in the secondary storage 12 and not in the primary storage 10.

Location 62: indicates the location of the file, which may comprise an address in the primary storage and/or secondary storage, such as the disk or tape volume and block address therein.

Stage Never Attribute 64: Comprises an attribute for the staging operation from the secondary storage 12 to the primary storage 10 for a file. The stage never attribute 64 indicates whether data transferred from the secondary storage 12 to the memory 14 for a particular file is to be destaged to the primary storage 10. If the stage attribute is set, then the data is read from the secondary storage into the memory 14 used by the application program 8 to buffer data and not staged to the primary storage 10, thereby not affecting the storage of other files in the primary storage. If the stage attribute not set, then the file is staged from the secondary storage 12 to memory 14, and then destaged from memory 14 to primary storage 10 according to the memory 14 caching algorithms. When the stage never attribute 64 is set, a window of data is staged into memory 14. The size of the window can be set by the user. If the file size is small, such as small database records, e.g., credit card transactions or other On-Line Transaction Processing (OLTP), then the user may set a small window size. If transferring large files, such as during a sequential access, then the window may be large.

09694078-062804

[0017] Further types of file metadata that may be included with the file metadata 50 are described in U.S. Patent No. 5,764,972, which was incorporated by reference above.

[0018] FIG. 3 illustrates how database tables and indexes 100 accessed by a database program 118 are comprised of multiple tablespaces 102a, b, c, wherein each tablespace
5 102a, b, c is comprised of multiple component data files 104a, b, c, 106a, b, c, and 108a, b, c that store the database data, such as database records.

[0019] In implementing the HSM features, the file system 4 would archive and release the files 104a, b, c, 106a, b, c, and 108a, b, c using the same criteria that is applied to determine regular files to release in the file system. Moreover the tablespace files 104a,
10 b, c, 106a, b, c, and 108a, b, c may be archived and released at different times, thereby leaving less than all the tablespace files in the primary storage 10. For instance, a more recently accessed tablespace file 104a, b, c, 106a, b, c, and 108a, b, c may remain in the primary storage 10 while a tablespace file that is one of the least recently used files may be marked for release.

15 [0020] In certain implementations, the partial release field 56 in the metadata 50 for each tablespace file would indicate a partial release, such that when the tablespace files 104a, b, c, 106a, b, c, and 108a, b, c are released a partial version of the tablespace file remains in the primary storage 10 that includes a first n bytes of the table space file, e.g., where n may be a predetermined or user specified number of bytes. FIG. 4 illustrates that
20 the primary storage 10 includes both complete tablespace files 104a, b, 106b, c, and 108c as well as partial versions 104c', 106a', 108a', and 108b' of certain files that have been archived in the secondary storage and released. For any file 104c, 106a, 108a, b released and having a partial version 104c', 106a', 108a', and 108b' remaining on the primary storage 10, one or more copies the full version of the file 104c, 106a, 108a, b are archived
25 on secondary storage as shown in FIG. 4.

[0021] FIGs. 5a, b illustrate logic implemented in the file system 4 to process an Input/Output (I/O) request, i.e., read or write, to a file, such as the tablespace files 104a, b, c, 106a, b, c, and 108a, b, c, in response to a read request received at block 200 from an

108290"062801

application program 8; such as the database program 118 that reads data from the memory 14. The logic of FIGs. 5a, b is invoked if the requested data at the offset location in the file is not in the memory 14. Otherwise, if the requested data is in the memory 14, then the data may be accessed directly from memory 14. If (at block 202) the metadata 50 for the file indicates that the target file is in the primary storage 10, which is indicated by the offline field 60 (FIG. 2) being set to "off", then the file system 4 accesses (at block 206) the requested offset location within the target file and reads (at block 208) a window length of data from the accessed location in the primary storage 10 until the window is filled in memory 14 or the end of the target file is reached. The window length may comprise a predefined or user specified amount of data that is read and stored in memory 14.

[0022] If (at block 204) the target file is offline and if (at block 210) the partial release attribute 56 (FIG. 2) is "off" (indicating that no partial version of the file is in the primary storage 10), then the file system 4 accesses (at block 212) the determined offset location in the target file in secondary storage 12 and reads (at block 214) the window length of data from the accessed location in the secondary storage 12 until the window is filled in memory 14 or the end of the target file is reached. After reading the window of data from the secondary storage 12 into memory 14, the file system 4 determines (at block 216) whether the stage never attribute 64 indicates that the data is not to be staged from the memory 14 to the primary storage 10. If the stage never attribute 64 indicates that data is not to be staged, then no action is taken to stage the data from the memory 14 to primary storage 10; otherwise, if the stage never attribute 64 indicates that data is to be staged, then the file system 4 would initiate action (at block 218) to stage the window of data in memory 14 to primary storage 10. In further implementations, the file system 4 may inhibit the staging of data to the primary storage 10 if the stage never attribute 64 indicates that data is not to be staged in alternative manners, such as by controlling the memory 14 manager.

09094078.062801

103290 " 82045850

[0023] If (at block 210) a partial version of the target file is in the primary storage 10, i.e., the first n bytes of the file are maintained on the primary storage 10, then the file system 4 determines (at block 220) whether the requested offset location is in the partial target file on the primary storage 10, i.e., within the first n bytes of the target file maintained in the partial target file. If not, control proceeds to block 212 to access the window of data at the offset location from the secondary storage 12. Otherwise, if the requested offset location is within the partial target file, then the file system 4 accesses (at block 222) the requested offset location in the partial file and reads the window length of data from the accessed location in the partial file. At block 224 in FIG. 5b, if the entire window length of data is read before reaching the end of the partial target file, then control ends; otherwise, the file system 4 determines (at block 226) the offset location in the target file on secondary storage 12 of the last read data at the end of the target partial file. The file system 4 then accesses (at block 228) the determined offset location of the target file in the secondary storage 12 and proceeds (at block 230) back to block 214 in FIG. 5a to continue reading data into the window in memory 14 from the secondary storage 12 until the window is filled or the end of the target file on secondary storage 12 is reached.

[0024] In further implementations, data being read into the window is immediately loaded into the memory 14 where it is available to application programs, such as the database program 118, even before the entire window has been filled. This feature allows the application program to obtain immediate access to data once the data is read into memory 14.

[0025] Still further, in performing an I/O request, the file system 4 may read multiple windows of data into memory 14 at a time, thereby prestaging data that may be requested. The number of windows of data to stage at a time in response to an I/O request may be a user settable parameter.

[0026] In implementations, where the application requesting data in files comprises a database program 118 requesting tablespace data in tablespace files 104a, b, c, 106a, b, c,

and 108a, b, c, the tablespace file data read from the secondary storage 12 is transferred to the area of memory 14 used by the database program 118. In this way, performance of the database program 118 is improved because data is made available to the database program 118 in memory 14 before the entire tablespace file 104a, b, c, 106a, b, c, and 108a, b, c is staged into primary storage 10. With the described implementations, data is provided directly from the secondary storage 12 to memory 14 and immediately made available in memory 14 to the database program 118.

[0027] Moreover, in implementations where the data transferred to memory 14 is not staged into the primary storage 10, providing the tablespace data to the database program 118 from the secondary storage 12 does not consume the primary storage 10 space and cause the file system 4 to release active data on the primary storage 10.

[0028] Further, in database implementations, the tablespace files 104a, b, c, 106a, b, c, and 108a, b, c may be released and a partial tablespace file 104b', 106a', 108a', 108b (FIG. 4) left on the primary storage 10 even while the tablespace 102a, b, c is in an open state, i.e., available to receive writes from the database program 118. This process of releasing tablespace files, including open tablespace files, according to migration criteria and leaving a partial tablespace file on primary storage 10 frees primary storage 10 space for more frequently accessed data from other applications or further tablespace files. Moreover, the described implementations provide fast access to tablespace files on the secondary storage by transferring windows of data to the memory to make data immediately available to the database program 118 before the entire tablespace file is staged.

Additional Implementation Details

[0029] The technique for managing data in a file system may be implemented as a method, apparatus or article of manufacture using standard programming and/or engineering techniques to produce software, firmware, hardware, or any combination thereof. The term "article of manufacture" as used herein refers to code or logic

09894078-062801

implemented in hardware logic (e.g., an integrated circuit chip, Field Programmable Gate Array (FPGA), Application Specific Integrated Circuit (ASIC), etc.) or a computer readable medium (e.g., magnetic storage medium (e.g., hard disk drives, floppy disks,, tape, etc.), optical storage (CD-ROMs, optical disks, etc.), volatile and non-volatile

5 memory devices (e.g., EEPROMs, ROMs, PROMs, RAMs, DRAMs, SRAMs, firmware, programmable logic, etc.). Code in the computer readable medium is accessed and executed by a processor. The code in which preferred embodiments of the configuration discovery tool are implemented may further be accessible through a transmission media or from a file server over a network. In such cases, the article of manufacture in which

10 the code is implemented may comprise a transmission media, such as a network transmission line, wireless transmission media, signals propagating through space, radio waves, infrared signals, etc. Of course, those skilled in the art will recognize that many modifications may be made to this configuration without departing from the scope of the present invention, and that the article of manufacture may comprise any information

15 bearing medium known in the art.

[0030] In the illustrations, a certain number of devices were shown. For instance, FIG. 1 illustrates one primary 10 and secondary 12 storage device. However, additional or fewer devices than shown may be used, e.g., more or less tape cartridges and tape drives may be included in the secondary storage 12. Further, the primary 10 and secondary 12

20 storage may be comprised of multiple storage devices and systems.

[0031] The described file management operations were performed by the file system component of an operating system. In alternative implementations, certain of the operations described as performed by the file system may be performed by some other program executing in the computer 2, such as an application program or middleware.

25 **[0032]** In certain described implementations, the files that were represented as partial files and read into the memory 14 from secondary storage 12 comprised tablespace files 104a, b, c, 106a, b, c, and 108a, b, c. Additionally, the above described technique for making tablespace data available to a database program 118 without having to stage in

09894078.062801

the entire data file is applicable to any data files used by any application program, and is not just limited to database programs.

[0033] In the described implementations, the primary storage 10 comprised a faster access storage than the secondary storage, and the storage media were different.

- 5 Alternatively, the primary storage 10 and secondary storage 12 may have the same access speeds and be implemented on the same storage media. Still further, the described file management technique may be used for files stored in one storage device that are not archived on a secondary storage.

- [0034] The program flow logic described in the flowcharts indicated certain events
10 occurring in a certain order. Those skilled in the art will recognize that the ordering of certain programming steps or program flow may be modified without affecting the overall operation performed by the preferred embodiment logic, and such modifications are in accordance with the preferred embodiments.

- [0035] The described implementations were discussed with respect to a Unix based
15 operating systems. However, the described implementations may apply to any operating system that provides file metadata and allows files in the system to be associated with different groups of users.

[0036] The described implementations may apply to situations where an application program accesses data maintained in multiple files.

- 20 [0037] In the described implementations, file information, such as the stage attributes and other file attributes was maintained in file metadata used by the file system. Alternatively, the file attribute information may be maintained in data structures and tables other than the file metadata used by the file system.

- [0038] The foregoing description of the preferred embodiments of the invention has
25 been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed. Many modifications and variations are possible in light of the above teaching. It is intended that the scope of the invention be limited not by this detailed description, but rather by the claims

1082908046850

appended hereto. The above specification, examples and data provide a complete description of the manufacture and use of the composition of the invention. Since many embodiments of the invention can be made without departing from the spirit and scope of the invention, the invention resides in the claims hereinafter appended.

5

SOLARIS is a trademark of Sun Microsystems, Inc.; UNIX is a registered trademark of The Open Group; SAM-FS is a trademark of LSC, Inc..

09894078 062801